心理科学进展 2018, Vol. 26, No. 10, 1765–1774
Advances in Psychological Science DOI: 10.3724/SP.J.1042.2018.01765

·研究前沿(Regular Articles)·

口语加工中的词语切分线索*

于文勃 梁丹丹

(南京师范大学文学院, 南京 210097)

摘 要 词是语言的基本结构单位, 对词语进行切分是语言加工的重要步骤。口语语流中的切分线索来自于语音、语义和语法三个方面。语音线索包括概率信息、音位配列规则和韵律信息, 韵律信息中还包括词重音、时长和音高等内容, 这些线索的使用在接触语言的早期阶段就逐渐被个体所掌握, 而且在不同的语言背景下有一定的特异性。语法和语义线索属于较高级的线索机制, 主要作用于词语切分过程的后期。后续研究应从语言的毕生发展和语言的特异性两个方面考察口语语言加工中的词语切分线索。

关键词 口语; 词语切分; 语音; 语义; 语法
分类号 B842

1 引言

语言单位包括语素、词、词组等, 其中, 词是最小的能独立运用的音义结合体, 是个体在头脑中存储的基本单位(张珊珊, 杨亦鸣, 2012)。口语语流是随时间变化的线性结构, 词语切分过程中, 词和词之间没有清晰可靠的边界, 不像文本阅读中有明确的空间线索(标点符号或空格), 早期的研究往往关注语义、语法等方面的线索信息, 但是婴幼儿在习得语言初期并不具有完备的语义知识和语法体系, 那么他们是如何进行切分的呢? 可以猜想, 语音信息可能是重要的线索。另一方面, 随着人工智能和语音合成等新技术的发展, 从语音层面探究词语的切分线索, 描绘人脑词语切分的内在过程俨然成为了当前心理学的研究热点。本文着重介绍口语加工中词语切分的语音线索, 随后介绍语法和语义线索, 最后对未来的研究提出一些建议。

2 词语切分中的语音线索

本部分内容聚焦词语切分的语音线索, 从概率信息、音位配列规则和韵律信息三个方面梳理相关研究。

2.1 概率信息

20 世纪 90 年代末, 研究者提出统计学习(statistical learning)的概念, 指个体自觉地运算刺激间的转换概率(transitional probability, TP)掌握统计规律的过程(Saffran, Aslin, & Newport, 1996; 唐溢等, 2015; Saffran & Kirkham, 2018), 这一认知过程也被认为是婴幼儿和成人在语流中切分词语、发现语法分类甚至是习得句法结构的重要方式(Newport, 2016)。

2.1.1 婴幼儿的研究

在口语语流中, 概率信息指单词内音节的转换概率高于单词间的音节, 比如词组 *pretty boy* 中, 音节 *re-ty* 间的转换概率要高于音节 *ty-boy* 间, 研究表明刚出生 8 个月的婴儿就已经具备了利用这一概率信息切分词语的能力(Aslin, Saffran, & Newport, 1998; Saffran, Aslin, et al., 1996; Saffran, Aslin, & Newport, 1996)。Saffran, Aslin 等(1996)设计了 4 个由 3 个音节组成的固定单词(*tupiro*, *golabu*, *bidaku*, *padoti*), 这些单词随机相连组成无意义音节串(*tupirogolabubidakupadotibidaku*……)。在完整单词 *tupiro* 中, 三个音节是固定连接的, 它们之间的转换概率为 1 (三个音节均为 *tupiro* 的内部音节, 同时出现), 但在跨界单词 *rogola* 中, 前两

收稿日期: 2017-12-27
*江苏高校优势学科建设工程资助项目(PAPD)资助。
通信作者: 梁丹丹, E-mail: ldd233@sina.com

个音节之间的转换概率为 0.33 (单词 *golabu* 可能出现在其他任意三个单词之后), 后两个音节的转换概率为 1, 因此跨界单词 *rogola* 中 *ro* 和 *go* 之间的转换概率较小, 意味着可能是词语边界, 所有无意义音节串均没有重音、停顿等线索, 只在转换概率上有所区分。实验分为学习阶段和测试阶段, 学习阶段让婴儿听 2 分钟的无意义音节串, 测试 1 发现婴儿对学习过的完整单词 *tupiro* 注视时间短, 对没学习过的单词 *tilado* 注视时间长; 测试 2 发现婴儿对学习过的完整单词 *tupiro* 注视时间短, 对学习过的跨界单词 *rogola* 注视时间长, 研究者认为这种去习惯化效应是因为婴儿以转换概率的高低划分词语边界, 对高转换概率的单词更为熟悉, 注视时间减少。

概率信息在词语切分中的作用也受到一些质疑, Estes (2012)认为大多数统计学习研究都是实验室研究, 而且实验材料为人工语法词, 这一学习机制是否能推广到自然语言环境中值得商榷; 另外, 也有研究者认为婴儿识别的音节串只是根据概率信息计算出来的音节单元, 并非是具有词汇属性的真实单词(Endress & Mehler, 2009; Perruchet & Poulin-Charronnat, 2012)。一些研究者通过实验在一定程度上反驳了以上质疑, 比如 Lew-Williams, Pelucchi 和 Saffran (2011)以意大利语为实验材料, 发现 8~10 个月的婴儿可以利用转换概率和词汇呈现形式来切分词语; Erickson, Thiessen 和 Estes (2014)发现 8 个月大的婴儿只会将高转换概率的音节串作为标签来对物体分类, 婴儿的这种分类能力被认为是基于真实词汇的, 因此研究者推断婴儿通过概率信息切分出来的音节串也具备一定的词汇属性。

2.1.2 成人的研究

相比于婴幼儿的研究, 成人的研究中更容易控制额外变量, 有助于深入分析概率信息在词语切分中的作用。Saffran, Aslin 等人(1996)的研究虽然证明婴儿可以通过音节间的概率信息切分口语语流, 但没有细致考察概率信息的载体。音节是我们直觉上最容易划分出来的最小语音单位, 一般以元音作为核心, 辅音在元音前面或后面, 共组成 4 种基本类型: (1)V, (2)C-V, (3)V-C, (4)C-V-C (林焘, 王理嘉, 2013), 那么概率信息的载体是元音、辅音还是整个音节这一问题并没有答案。近年来以成人为被试的研究发现, 不同语言背景下个体对承载概率信息的语音载体有着不同的偏好(Bonatti, Peña, Nespor, & Mehler, 2005; Gómez, Mok, Ordin, Mehler, & Nespor, 2017)。Bonatti 等人(2005)在经典的转换概率范式基础上, 分别在元音和辅音层面上控制音节间的概率信息, 结果发现当辅音为载体时, 法语被试能够更好地利用概率信息进行词语切分, 研究者认为这是因为在印欧语系中辅音对单词识别的作用大于元音。Gómez 等人(2017)以粤语母语者为被试, 沿用了 Bonatti 等(2005)的实验范式, 首先在材料中保证了音节间的转换概率恒定(音节 *ge* 后接音节 *du* 或 *dy*), 然后分别改变元音间的概率信息(含元音/*u*/的音节后接含元音/*e*/的音节的概率为 0.75, 接含元音/*i*/的音节的概率为 0.25)和辅音间的概率信息(含辅音/*h*/的音节后接含辅音/*b*/的音节的概率为 0.75, 接含辅音/*g*/的音节的概率为 0.25), 结果发现, 相比于辅音条件, 粤语母语者在元音条件下能更好地利用概率信息切分词语。不同于大部分印欧语系语言, 以汉语普通话、粤语和越南语等为代表的汉藏语系语言具有声调这一超音段特征, Gómez 等人(2017)还发现随着声调信息的加入粤语被试对词语切分的准确率进一步提高。可见, 虽然利用概率信息切分词语是人类普遍的能力, 但在不同语言背景中表现形式并不相同。

2.2 音位配列规则

每种语言都有自己的语音音位配列规则(phonotactics), 符合配列规则的音位搭配出现频率高, 不符合的出现频率低甚至为 0, 比如在英语中/*mk*/就是高频辅音搭配, 而/ *ŋk* /是低频辅音搭配。当个体在语流中识别到不可能同处于一个音节的两个音位时, 会倾向认为二者之间存在音节边界, 而如果前后两个音节分别是单音节词, 那么在切分音节的同时就完成了词语的切分(McQueen, 1998; Suomi, McQueen, & Cutler, 1997; Tremblay & Spinelli, 2013)。在荷兰语的研究中, McQueen (1998)采用词语指认范式, 要求被试在听到无意义双音节中的真词时迅速报告, 比如在无意义双音节词 *pil.vrem* 和 *pilv.rem* 中, 真词音节均为 *pil*, 但是前者辅音/*l*/和/*v*/分别处在两个音节中, 后者辅音/*l*/和/*v*/处在同一个音节内。结果发现, 被试在第一种条件下报告真词的反应时更短, 准确率更高, 研究者指出在荷兰语中辅音/*l*/和/*v*/不能处于同一音节内, 与第一种条件刺激的发音方式相匹

配, 被试在听到双音节词时更容易判断两个音位之间有音节边界, 进而完成了对真词的切分。

紧张性和松弛性是普遍存在的音位对立特征, 既可以表现在元音上, 也可以表现在辅音上, 紧元音(tense vowel)听起来强而长, 松元音(lax vowel)听起来短而弱(王理嘉, 1991)。在英语中, 紧元音(如/u/、/i:/)可以作为词尾音, 而松元音(如/i/、/ʌ/)不可以, Skoruppa, Nevins, Gillard 和 Rosen (2015)发现在语音片段 *this is a /naɪzʌteɪ/*中, 个体倾向将其切分成 *this is a /naɪ/ /zʌteɪ/*而不是 *this is a /naɪzʌ/ /teɪ/*, 这说明元音的松紧性提供了必要的线索。音位配列规则可以看作是音位间、音节间概率信息的延伸, 暴露在语言环境下的个体可以通过它们之间的概率信息掌握音节与音节之间、词与词之间的边界, 进而内化为语音规则, 而无需特定的习得过程。

2.3　韵律信息

语言的语音结构由音段结构和超音段结构两部分组成(何善芬, 1989), 音段结构就是上文提到的音节, 也指其内部的元音和辅音, 词语切分中的概率信息和音位配列规则主要作用在音段结构上; 超音段特征包括音高、强度以及时间特性, 由音位或音位群负载(杨玉芳, 黄贤军, 高路, 2006), 相关的研究表明, 多种超音段信息也可以作为线索帮助个体切分口语语流。在韵律音系学中, 韵律特征(语调、时域分布和重音)主要通过超音段特征实现, 因此本部分所介绍的超音段信息也可以被称作韵律信息。

2.3.1　词重音

一段语流中各音节声音响亮程度并不完全相等, 在语流中听起来比其他音节突显的音节称为重音音节。重音可以分成词汇层面的词重音和句子层面的句重音或重读。词重音有词汇属性, 具有语法和词汇意义, 起到辨义作用, 而句重音彰显话语组织的突出焦点, 具有语用功能(何善芬, 1989; 许希明, 沈家煊, 2016)。Hyman (2009)将世界语言划分为重音语言和声调语言, 前者以英语为代表, 带有词层面的节律特征, 后者以汉语普通话为代表, 带有词层面的音高特征。相关的研究表明, 以重音语言为母语的个体能够利用词重音作为线索切分语流。

英语是自由重音语言, 单音节词不会遇到重音分配的问题, 多音节词的重音分配位置不固定, 虽然大多数单词词重音位于第一音节(如 *melody, polish, favorite*)但也可能位于其他音节上(如 *begin, anecdotal*)。Cutler 和 Carter (1987)通过语料库调查发现在英语的实义词中, 强音节开头的数量是弱音节开头数量的三倍, 而且前者出现的频率也是后者的两倍, 因此他们推断英语母语者会通过词重音确定词语的起始位置。Cutler 和 Norris (1988)设计了两类无意义音节: *mintayve* 和 *mintesh*, 前者由两个完整元音音节组成, 记为 SS (强强)音节, 后者由一个完整元音音节和一个半元音音节组成, 记为 SW (强弱)音节, 实验要求被试在听无意义音节的同时检测真词(如 *mint*)的出现, 结果发现被试对 SS 音节中真词的反应时间显著长于 SW 音节, 这可能是因为音节 *mint* 和 *tayve* 均是重音音节, 二者会竞争辅音/t/, 进而干扰对真词 *mint* 的识别, 而 SW 音节中不存在竞争关系。婴儿的研究也证实了词重音作为线索对切分词语的作用, Jusczyk, Houston 和 Newsome (1999)采用转头偏好范式, 考察 7.5 个月婴儿的音节识别能力, 结果发现他们对符合英语词重音模式(重音为第一音节)的双音节单词有偏好, 而对于不符合词重音模式的单词没有偏好。

虽然词重音可以作为英语词语切分的线索, 但是这一线索并非具有跨语言的普遍性。法语词重音形式与英语不同, 所有词重音均在词末音节上(林焘, 王理嘉, 2013), 属于固定重音语言, 研究发现法语母语者并非通过重音而是通过音节的完整性来切分词语(Mehler, Dommergues, Frauenfelder, & Segui, 1981); 而在同样是重音语言的西班牙语中, 母语者在切分词语过程中会结合音节的数量和重音两方面线索(LaCross et al., 2016)。

2.3.2　音高和时长信息

韵律结构普遍存在于所有语言中, 每一个韵律结构都会存在韵律边界, 通常伴随语段末音段延长、无声段以及相对较大的音高移动(李卫君, 杨玉芳, 2010)。研究指出这些音高和时长变化在语音歧义词的切分过程中起着消解歧义的作用(Christophe, Peperkamp, Pallier, Block, & Mehler, 2004; Gout, Christophe, & Morgan, 2004; Shatzman & McQueen, 2006)。在 Christophe 等人(2004)以法语为材料的实验中, 目标词可以和后面单词的首音节(歧义音节)组成合乎语义的竞争词, 但是目标词(*chat*)和歧义音节(*gri*)或者处在韵律短语内

部(如[*d'un chat grincheux*]), 或者处在韵律短语边界处(如[*le gros chat*] [*grimpait*...])。他们发现被试对目标词的反应情况受到韵律边界的调节, 如果目标词和歧义音节分属于不同的韵律短语, 那么韵律边界有助于切分二者, 避免形成竞争词干扰目标词的识别。

韵律边界对词语的切分体现在音高和时长两方面信息的共同作用上, 那么两者中单独一个因素是否也能够起到切分词语的作用呢? Shatzman 和 McQueen (2006)采用跨通道语义启动范式考察荷兰语 *Ze beeft wel eens pot gezegd* 中辅音/*s*/的时长对歧义词组的切分影响(*eens pot*, *een spot*)。结果发现, 当辅音/*s*/持续时间较短时, 被试更早地对目标词(*pot*)对应的图片进行注视, 这是因为位于词尾(*eens*)的辅音/*s*/的时长要短于位于词首(*spot*)的情况, 因此被试将较短的/*s*/切分成前一个单词的词尾, 进而对目标词(*pot*)加工更快。除此之外, 关于抑扬−扬抑规律(ITL, Iambic\Trochaic law)的研究也提供了音高和时长信息是如何在词语切分中起线索作用的证据(Frost, Monaghan, & Tatsumi, 2017; Langus et al., 2016)。早在一百多年前, 研究者就发现个体具有根据强度、时长和音高等声学特征将声音序列进行组块化的倾向(Bolton, 1894; Woodrow, 1909)。Hayes (1995)提出节奏感知的抑扬−扬抑规律(Iambic\Trochaic Law): 在强度参数上, 个体对节奏感知有强弱形式的扬抑偏好(后续研究发现音高参数与强度参数规律相同); 在时长参数上, 个体对节奏有短长形式的抑扬偏好; 作者进一步指出这一规律不仅仅是语言的结构形式, 也是个体组织、切分语言的方式。近年来的实证研究将焦点放在抑扬−扬抑规律对词语切分作用的跨语言特性上。Langus 等人(2016)以意大利语、土耳其语和波斯语母语者为被试, 以重复出现、顺序固定的无意义音节为材料(*pa su tu ke ma vi bu go ne du*), 每隔一个音节改变音节的时长(180~400 ms)或基频 F0 (180~400 Hz), 熟悉阶段要求被试认真听语音材料, 测试阶段给被试呈现音节对, 如 *pa-su*, 要求判断其是否刚刚出现过。结果发现在音高参数上, 三组被试成绩相当且正确率较高(0.7~0.8), 说明他们都以扬抑形式切分音节, 能够区分音节对 *pa-su* 和 *su-tu*; 但是在时长参数上, 意大利母语者判断的正确率显著高于其他两组被试, 说明只有意大利母语者能够利用时长线索正确切分音节, 即词语切分过程受到语言经验的影响。不过, Frost 等人(2017)的研究与此结论完全相反, 他们考察了日语母语者和英语母语者, 发现在时长参数下, 两组被试的回答正确率相当, 研究者认为抑扬−扬抑规律对词语的切分效应是一般性的认知机制, 具有跨语言的普遍性。虽然两个研究采用的实验范式相同, 自变量和因变量指标也基本一致, 但两者在材料设置上有细微差别, 前者的音节呈现顺序固定, 后者的音节呈现顺序随机变化, 而且后者的作答形式为迫选, 要求被试在两个音节对中选择更像单词的一个, 这可能是造成两个研究结果相悖的原因。总之, 关于时长、音高等声学信息在词语切分中作用的研究刚刚起步, 在研究范式和材料上都有不完善的地方, 还需要更多的研究加以对比。

2.3.3 韵律特征的规律性

上文介绍的线索信息在词语切分过程中大多作用在目标词附近, 可以看作是个体利用即时信息对口语语流进行切分, 但也有研究发现当个体对语流进行加工时, 如果前段语流的韵律特征(时长、音高)呈规律性的变化, 那么个体会以相同的变化模式切分后续语流(Brown, Dilley, & Tanenhaus, 2012; Brown, Salverda, Dilley, & Tanenhaus, 2015; Dilley, & McAuley, 2008; Dilley, Mattys, & Vinke, 2010)。

Dilley 和 McAuley (2008)设计了一系列由 8 个音节组成的单词串(*skirmish princess side kick stand still*), 其中前两个单词为重音在第一音节的双音节单词, 后面 4 个单音节单词可以组成多种音节形式的单词(*sidekick standstill, side kickstand still*)。考察音高线索时, 将前两个单词(*skirmish* 和 *princess*)的基频 F0 设置成由高到低(270~280 Hz 到 170~180 Hz)或相反的变化趋势(如图 1), 其中单音节条件中(图 1 第一行)第五个音节(*side*)的 F0 由高到低(270~280 Hz 到 170~180 Hz), 双音节条件中(图 1 第二行)第五个音节(*side*)的 F0 为低(170~180 Hz), 两种条件下最后三个音节的 F0 保持一致。实验任务要求被试在听到单词串后报告他们听到的最后一个单词, 结果发现在单音节条件下, 被试会按照"高低高低"的组合规律切分单词, 将 *kick* 和 *stand* 听成一个合成词 *kickstand*, 最后报告单音节单词 *still*; 然而在双音节条件下,

被试更多地报告双音节单词 *standstill*。

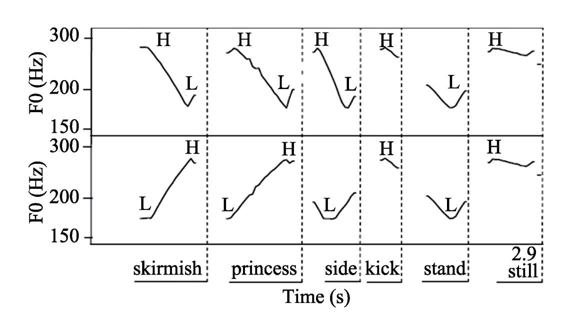


图 1　实验材料示意图
(资料来源：Dilley & McAuley, 2008)

语速通常被定义为单位时间内听到的音段或音节的数量(Reinisch, 2016), 语速快意味着每个音节的时长短, 语速慢意味着每个音节的时长长, 因此, 语速可以被看作是音节时长的一种表现形式。相关的研究表明, 语速也会对目标词的切分产生影响(Baese-Berk et al., 2014; Dilley & Pitt, 2010; Morrill, Baese-Berk, Heffner, & Dilley, 2015; Morrill, Dilley, McAuley, & Pitt, 2014)。在句子 *Deena doesn't have any leisure or time* 中, 通过 PSOLA 软件调整语句前段音节串的语速(正常语速, 1.9 倍正常语速和 0.6 倍正常语速), 结果发现当目标词 *or* 前面的单词语速较慢时, 被试倾向于报告没有听到目标词 *or* (*leisure time*), 而语速较快时, 被试倾向于报告听到目标词 *or* (*leisure or time*)。研究者推断较慢的语速中, 被试期待音节的时长较长, 单词 *leisure* 与 *or* 协同发音导致被试感知不到目标词 *or* (Dilley & Pitt, 2010)。在跨语言的研究中, Lai 和 Dilly (2016)采用相同的实验范式, 发现在汉语语句*我们想摆一盆花在窗台上*中, 音节 *yi1* 的识别也受到语速的影响; 而且即使过滤掉语义信息只保留基频信息, 目标词远端的韵律特征依旧可以影响词语切分的结果(Dilley et al., 2010)。

口语词语切分过程中, 语音范畴提供的线索大致可以分为音段线索和韵律线索两类, 虽然线索载体和作用方式都不同, 但是两类线索均是个体在接触语言早期就能够习得的, 尤其是已有研究证实个体在 1 岁以内就可以利用概率信息和词重音信息切分词语; 尽管抑扬-扬抑规律对词语切分的线索作用只得到成人研究的支持, 但是相关研究已经表明个体在出生伊始就具备抑扬或扬抑偏好(Abboub, Nazzi, & Gervain, 2016), 可以设想婴幼儿在切分词语时会综合使用音段和韵律线索。另外, 语音切分是语音合成的逆向过程, 现有的研究成果可以为增强合成语音的表现力和自然度提供帮助(李勇, 魏玮, 王柳渝, 2017)。

3　词语切分中的语法和语义线索

相比于语音线索, 语法和语义层面的信息对词语切分影响的研究较少, 而且考虑到这二者均是较高级的语言知识, 因此研究对象主要是成人。

3.1　语法线索

Cole, Jakimik 和 Cooper (1980)采用错误发音听辨任务(listening for mispronunciations task)考察语法线索在词语切分中的作用, 发现语流前段的语法结构能够提示被试后面的词语形式, 比如 *he just hated* 结构提示被试后面为 *doing* 形式, 因此在听到错误辅音/*k*/ (正确形式为 *forgetting*, 错误形式为 *forketting*)时, 被试可以迅速报告错误发音, 而 *he was noted* 结构提示被试后面为 *for doing* 形式, 因此被试听到错误辅音/*k*/时, 还要进一步排除其他可能。Mattys, Melhorn 和 White (2007)考察了英语主谓一致原则对语音歧义词切分的影响, 实验包括 3 种听觉材料, 中性词组(*take spins*, *takes pins*)、单数主语句子(*That woman takes pins*, *That woman take spins*)和复数主语句子(*Those women takes pins*, *Those women take spins*), 在每一个试次中, 先呈现视觉目标词(*spins* 或 *pins*), 随后呈现听觉词组或句子, 要求被试尽快判断听觉材料中是否包括视觉目标词。结果发现, 在中性词组条件中出现了一致性效应, 即当目标词与听觉刺激一致时(*spins* 和 *take spins*, *pins* 和 *takes pins*), 被试的反应时要短于二者不一致时(*spins* 和 *takes pins*, *pins* 和 *take spins*); 但在复数主语句子条件下, 被试反应的一致性效应消失, 研究者认为复数条件下句子的主语 *Those women* 导致被试期待听觉刺激 *take spins* 的出现, 因此即使目标词 *pins* 和语句中的单词 *Those women takes pins* 一致, 对目标词的切分也会受到语法的抑制。

近年来, 一些研究者试图从脑神经活动的角度揭示个体切分词语的内在机制。Ding, Lucia, Zhang, Tian 和 Poeppel (2016)采用脑磁图技术, 向被试呈现没有韵律信息且均由单音节单词组成的

中英文句子(*历史不会重演, white cars need gas*), 结果发现当被试听到符合语法结构的词组(*历史, white car*)或者句子(*历史不会重演*)时, 大脑皮层会出现较明显的电磁频谱反应, 研究者将其称为大脑皮层的“神经锁定” (neuro entrainment)现象, 借于此他们推断个体能够以语法知识切分语流, 并建构语义表征。

3.2 语义线索

索绪尔(De Saussure & Baskin, 1916)在《普通语言学教程》书中指出要通过音节的意义对语流进行切分, 从而保证被切分的音节有对应的实体, 比如, 法语音节串 *sižlaprà* 只能切分成 *si-ž-la-prà* (如果我拿走它)和 *si-ž-l-aprà* (如果我掌握它)两种。虽然索绪尔的设想较为粗糙, 没有得到实验证据的支持, 但随着语言知识的增长, 个体对词语的切分必然受到语义的限制。Norris, McQueen 和 Culter (1995)提出口语词语切分的可能性限制原则(Possible-Word Constraint, PWC), 认为在口语词语加工中个体头脑的候选词语必须能够解释语流中的所有音位, 只有这样才能完成词语识别, 进而完成词语切分。在研究中, Norris, McQueen, Cutler 和 Butterfield (1997)要求被试在听到音节串的同时识别真词, 结果发现在音节串 *fapple* 中对真词 *apple* 的识别比在音节串 *vuffapple* 中更加困难, 这是因为音位/*f*/无法单独构成一个单词, 不利于切分音节串, 而 *vuff* 构成单词的可能性较大, 有利于切分音节串。

近年来, 视听跨通道词语启动范式普遍被使用在词语切分的研究中(White, Mattys, & Wiget, 2012)。White 等人(2012)设计了强语义关联和弱语义关联的词组(*oil tanker* 和 *seal tanker*), 并从模拟对话中切分出真实词组作为实验材料。实验过程中, 首先呈现听觉词组作为启动刺激, 随后呈现视觉目标词, 要求被试判断目标词是否为真实单词, 视觉目标词包括三种情况: 与探测词组的结尾词相同、与探测词组无关和非词。结果发现在强语义相关的探测条件下被试对与探测词组结尾相同的视觉单词判断更快, 这说明强语义探测刺激具有启动效应, 加快了对目标词 *tanker* 的切分。

4 多种线索的交互作用

在实际的口语加工中, 切分词语是个极其复杂的过程, 受到多种线索的协同(竞争)作用, 一些研究考察了韵律特征(重音)、音位规则、语义和语法等线索的相对权重(Babineau, Shi, & Achim, 2017; Heffner, Dilley, McAuley, & Pitt, 2013; Mattys, 2004)。在法语连音(French liaison)的研究中, Babineau 等(2017)发现句法规则对连音的切分作用最大, 而语音线索只起辅助作用; 另一方面, 环境背景和被试的策略也会影响词语切分过程(Mattys, White, & Melhorn, 2005; Morrill et al., 2015)。Mattys 等人(2005)的研究结果发现, 在安静环境下语义的影响权重最大, 随后是词汇信息和音位规则, 而重音的线索作用最小, 但在噪声环境中韵律特征等低层级线索的作用变大。

通过行为学实验可以判断个体对不同线索的依仗程度, 但是不能探究个体利用多种线索的时间进程, 大量事件相关电位的研究证实在词语切分过程中, 语音线索的作用是即时的, 不受高级线索的影响, 而语义、语法线索的作用时间主要位于整合语义的最后阶段(Steinhauer, Alter, & Friederici, 1999; 张辉, 孙和涛, 顾介鑫, 2013)。在 Steinhauer 等(1999)以德语为材料的研究中, 句子的每一个韵律短语边界都会引起被试顶叶脑区的活动, 出现中止正漂移的脑电成分(closure positive shift, CPS), 而如果韵律线索切分的句子结构与句法结构矛盾, 还会出现一个双向的 N400-P600 成分(biphasic N400-P600)。张辉等人(2013)以相同的实验范式考察汉语母语者对四字成语材料的切分情况, 实验过程中向被试呈现两种朗读模式的成语(2+2, 1+3), 其中每种朗读模式中一半是符合成语句法结构的(*废寝/忘食, 狐/假虎威*), 一半是不符合的(*恩重/如山, 双/喜临门*)。结果发现韵律节奏主效应显著, 无论材料是否符合句法模式, 只要以“1+3”节奏朗读时, 都会激发被试双向的 N400-P600 成分, 而以“2+2”节奏朗读则不会出现此成分。这说明个体在运用韵律信息切分语流时有一定的独立性, 切分早期并不受到语义、语法等高级线索的干扰。

5 小结与展望

本文系统地梳理了语音、语法和语义线索对口语词语切分的作用, 近年来的研究一方面集中在婴幼儿词语切分的线索机制上, 另一方面聚焦于成人是如何综合利用多种线索进行词语切分的, 笔者认为目前的研究仍有不足之处, 可以从以下

两个方面丰富、扩展。

5.1　从毕生发展角度考察词语切分线索

语言发展会历经个体从出生到死亡的所有阶段, 目前的研究主要集中在成人口语词语切分上, 婴幼儿的研究才刚刚起步。一直以来, 婴幼儿是如何掌握词语这一问题始终困扰心理学家和语言学家, 除了本文提到的概率信息外是否还有其他线索呢? 一些研究发现婴幼儿对词语的习得受到他们接触词语的频率(Ambridge, Kidd, Rowland, & Theakston, 2015)、时间分布、空间分布和文本环境(Roy, Frank, DeCamp, Miller, & Roy, 2015)的影响; 语料库的调查也发现婴儿所接触的语料中有 9%是单个单词(single word), 出现频率较高的是 *come*、*go*、*up* 和 *down* 等(Ambridge & Lieven, 2011), 因此可以猜想婴儿首先要掌握单个单词, 随后以此作为"据点"切分词组和句子, 进而掌握新词, 但这一假设仍需要更多实验证据的支持。除此之外, 语言加工的老化研究也是近期兴起的热点问题, 词语切分能力是否与语义加工一样存在老化现象呢? 如果存在, 是单一线索使用能力下降还是多种线索使用能力共同下降呢?

5.2　从语言特异性角度考察词语切分线索

跨语言的研究已经证实, 个体在切分语言的过程中受到母语语音结构的影响(Cutler & Otake, 1994; Mehler et al., 1981; LaCross et al., 2016)。印欧语系中, 语言中的一个基本单位(词)通常对应若干音节, 因此对词语的切分首先要解决的问题是一个单词对应几个音节; 但汉语的音节结构有其独特的地方, 首先, 每一个音节有一个声调, 使同一个音节中的各个音位有一种向心力, 内聚为一个整体, 从而能清楚地与其他音节区别开来(徐通锵, 2001); 此外, 汉语音节以元音结尾占多数, 辅音结尾的只有/*n*/和/*ng*/两种, 以元音开头的音节又极少, 这都大大降低了连读的可能性; 最后, 按照徐通锵(2010)的说法, 汉语的特点是"1 个字·1 个音节·1 个概念", 英语中相应的结构是"1 个词·n 个音节·1 个概念", 即汉语音节与意义是一一对应的关系, 可见在汉语中识别音节并不存在困难。不过, 汉语词汇化过程中表现出明显的双音化倾向, 冯胜利(1998)也指出由于自然音步的影响, 汉语普通话母语者习惯使用双音节词, 端木三(2000)也以*小火/车*为例子, 指出汉语中的韵律结构对句法结构具有一定的限制作用, 那么这种双音节倾向是否是汉语母语者在语音层面上切分语流的线索呢? 这有待于进一步考证。另一方面, 汉语没有明显的语法形态, 同音字数量多, 这导致了汉语中存在大量的同音异构形式, 比如*炒饭*这一结构, 既可以表示动宾含义炒饭这一动作, 也可以表示偏正含义炒饭这一实物, 再比如结构*小张师傅手艺很好*也存在歧义, 可以表示小张师傅本人, 也可以指代小张的师傅, 今后的研究可以发掘韵律信息在此类结构中的切分作用。

词语切分是语言加工研究的根本问题, 口语状态下的词语切分是自然交际中的关键环节, 未来应更广泛地从不同线索、不同视角、不同语言展开对这一问题的探讨, 不但可以揭示出某种具体语言中口语加工时的词语切分过程, 而且可以在此基础上得出具有普遍性的口语词语切分模型。

参考文献

端木三. (2000). 汉语的节奏. *当代语言学,* (4), 203−209.

冯胜利. (1998). 论汉语的 "自然音步". *中国语文,* (1), 40−47.

何善芬. (1989). 英语超音段音位及其辨义功能. *外国语,* (6), 66−69.

李勇, 魏玙, 王柳渝. (2017). 基于 PSOLA 与 DCT 的情感语音合成方法. *计算机工程,* (12), 278−282.

李卫君, 杨玉芳. (2010). 绝句韵律边界的认知加工及其脑电效应. *心理学报, 42*(11), 1021−1032.

唐溢, 张智君, 曾玫媚, 黄可, 刘炜, 赵亚军. (2015). 基于名人面孔视觉特征和语义信息的视觉统计学习. *心理学报, 47*(7), 837−850.

王理嘉. (1991). *音系学基础.* 北京: 语文出版社.

林焘, 王理嘉. (2013). 语音学教程. 北京: 北京大学出版社.

徐通锵. (2001). *基础语言学教程.* 北京: 北京大学出版社.

徐通锵. (2010). *徐通锵文选.* 北京: 北京大学出版社.

许希明, 沈家煊. (2016). 英汉语重音的音系差异. *外语教学与研究, 49*(5), 643−656.

杨玉芳, 黄贤军, 高路. (2006). 韵律特征研究. *心理科学进展, 14*(4), 546−550.

张辉, 孙和涛, 顾介鑫. (2013). 成语加工中韵律与句法互动的事件相关电位研究. *外国语: 上海外国语大学学报,* (1), 22−31.

张珊珊, 杨亦鸣. (2012). 从记忆编码加工看人脑中的基本语言单位——一项基于单音节语言单位的 ERPs 研究. *外语与外语教学,* (2), 1−6.

Abboub, N., Nazzi, T., & Gervain, J. (2016). Prosodic grouping at birth. *Brain and Language, 162*, 46−59.

Ambridge, B., Kidd, E., Rowland, C. F., & Theakston, A. L. (2015). The ubiquity of frequency effects in first language

acquisition. *Journal of Child Language, 42*(2), 239−273.

Ambridge, B., & Lieven, E. V. M. (2011). *Child language acquisition: Contrasting theoretical approaches*. Cambridge: Cambridge University Press.

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science, 9*(4), 321−324.

Babineau, M., Shi, R., & Achim, A. (2017). Contextual factors in lexical processing: The case of French Liaison. *Language, Cognition and Neuroscience, 32*(4), 457−470.

Baese-Berk, M. M., Heffner, C. C., Dilley, L. C., Pitt, M. A., Morrill, T. H., & McAuley, J. D. (2014). Long-term temporal tracking of speech rate affects spoken-word recognition. *Psychological Science, 25*(8), 1546−1553.

Bolton, T. L. (1894). Rhythm. *The American Journal of Psychology, 6*(2), 145−238.

Bonatti, L. L., Peña, M., Nespor, M., & Mehler, J. (2005). Linguistic constraints on statistical computations: The role of consonants and vowels in continuous speech processing. *Psychological Science, 16*(6), 451−459.

Brown, M., Dilley, L. C., & Tanenhaus, M. K. (2012, January). *Real-time expectations based on context speech rate can cause words to appear or disappear*. Proceedings of the Annual Meeting of the Cognitive Science Society. Austion, TX.

Brown, M., Salverda, A. P., Dilley, L. C., & Tanenhaus, M. K. (2015). Metrical expectations from preceding prosody influence perception of lexical stress. *Journal of Experimental Psychology: Human Perception and Performance, 41*(2), 306−323.

Christophe, A., Peperkamp, S., Pallier, C., Block, E., & Mehler, J. (2004). Phonological phrase boundaries constrain lexical access I. Adult data. *Journal of Memory and Language, 51*(4), 523−547.

Cole, R. A., Jakimik, J., & Cooper, W. E. (1980). Segmenting speech into words. *The Journal of the Acoustical Society of America, 67*(4), 1323−1332.

Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech & Language, 2*(3−4), 133−142.

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 14*(1), 113−121.

Cutler, A., & Otake, T. (1994). Mora or phoneme? Further evidence for language-specific listening. *Journal of Memory and Language, 33*(6), 824−844.

De Saussure, F., & Baskin, W. (1916). *Course in general linguistics*. London: Duckworth.

Dilley, L. C., & McAuley, J. D. (2008). Distal prosodic context affects word segmentation and lexical processing. *Journal of Memory and Language, 59*(3), 294−311.

Dilley, L. C., Mattys, S. L., & Vinke, L. (2010). Potent prosody: Comparing the effects of distal prosody, proximal prosody, and semantic context on word segmentation. *Journal of Memory and Language, 63*(3), 274−294.

Dilley, L. C., & Pitt, M. A. (2010). Altering context speech rate can cause words to appear or disappear. *Psychological Science, 21*(11), 1664−1670.

Ding, N., Lucia, M., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience, 19*(1), 158−164.

Endress, A. D., & Mehler, J. (2009). The surprising power of statistical learning: When fragment knowledge leads to false memories of unheard words. *Journal of Memory and Language, 60*(3), 351−367.

Estes, K. G. (2012). Infants generalize representations of statistically segmented words. *Frontiers in Psychology, 3*(3), 447.

Erickson, L. C., Thiessen, E. D., & Estes, K. G. (2014). Statistically coherent labels facilitate categorization in 8-month-olds. *Journal of Memory and Language, 72*, 49−58.

Frost, R. L. A., Monaghan, P., & Tatsumi, T. (2017). Domain-general mechanisms for speech segmentation: The role of duration information in language learning. *Journal of Experimental Psychology Human Perception and Performance, 43*(3), 466−476.

Gómez, D. M., Mok, P., Ordin, M., Mehler, J., & Nespor, M. (2017). Statistical speech segmentation in tone languages: The role of lexical tones. *Language and Speech, 61*(1), 84−96.

Gout, A., Christophe, A., & Morgan, J. L. (2004). Phonological phrase boundaries constrain lexical access II. Infant data. *Journal of Memory and Language, 51*(4), 548−567.

Hayes, B. (1995). *Metrical stress theory: Principles and case studies*. Chicago: University of Chicago Press.

Heffner, C. C., Dilley, L. C., McAuley, J. D., & Pitt, M. A. (2013). When cues combine: How distal and proximal acoustic cues are integrated in word segmentation. *Language and Cognitive Processes, 28*(9), 1275−1302.

Hyman, L. M. (2009). How (not) to do phonological typology: the case of pitch-accent. *Language Sciences, 31*(2−3), 213−238.

Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology, 39*(3), 159−207.

LaCross, A., Liss, J., Barragan, B., Adams, A., Berisha, V., McAuliffe, M., & Fromont, R. (2016). The role of stress

and word size in Spanish speech segmentation. *The Journal of the Acoustical Society of America, 140*(6), EL484−EL490.

Lai, W., & Dilley, L. (2016). *Cross-linguistic generalization of the distal rate effect: Speech rate in context affects whether listeners hear a function word in Chinese Mandarin*. Proceedings of 2016 Speech Prosody, Boston, MA.

Langus, A., Seyed-Allaei, S., Uysal, E., Pirmoradian, S., Marino, C., Asaadi, S., ... Nespor, M. (2016). Listening natively across perceptual domains?. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(7), 1127−1139.

Lew‐Williams, C., Pelucchi, B., & Saffran, J. R. (2011). Isolated words enhance statistical language learning in infancy. *Developmental Science, 14*(6), 1323−1329.

Mattys, S. L. (2004). Stress versus coarticulation: Toward an integrated approach to explicit speech segmentation. *Journal of Experimental Psychology: Human Perception and Performance, 30*(2), 397−408.

Mattys, S. L., Melhorn, J. F., & White, L. (2007). Effects of syntactic expectations on speech segmentation. *Journal of Experimental Psychology: Human Perception and Performance, 33*(4), 960−977.

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General, 134*(4), 477−500.

McQueen, J. M. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory and Language, 39*(1), 21−46.

Newport, E. L. (2016). Statistical language learning: Computational, maturational, and linguistic constraints. *Language and Cognition, 8*(3), 447−461.

Mehler, J., Dommergues, J. Y., Frauenfelder, U., & Segui, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior, 20*(3), 298−305.

Morrill, T. H., Dilley, L. C., McAuley, J. D., & Pitt, M. A. (2014). Distal rhythm influences whether or not listeners hear a word in continuous speech: Support for a perceptual grouping hypothesis. *Cognition, 131*(1), 69−74.

Morrill, T., Baese-Berk, M., Heffner, C., & Dilley, L. (2015). Interactions between distal speech rate, linguistic knowledge, and speech environment. *Psychonomic Bulletin & Review, 22*(5), 1451−1457.

Norris, D., Mcqueen, J. M., & Cutler, A. (1995). Competition and segmentation in spoken-word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(5), 1209−1228.

Norris, D., McQueen, J. M., Cutler, A., & Butterfield, S. (1997). The possible-word constraint in the segmentation of continuous speech. *Cognitive Psychology, 34*(3), 191−243.

Perruchet, P., & Poulin-Charronnat, B. (2012). Beyond transitional probability computations: Extracting word-like units when only statistical information is available. *Journal of Memory and Language, 66*(4), 807−818.

Reinisch, E. (2016). Natural fast speech is perceived as faster than linearly time-compressed speech. *Attention, Perception, & Psychophysics, 78*(4), 1203−1217.

Roy, B. C., Frank, M. C., DeCamp, P., Miller, M., & Roy, D. (2015). Predicting the birth of a spoken word. *Proceedings of the National Academy of Sciences of the United of America, 112*(41), 12663−12668.

Saffran, J. R., & Kirkham, N. Z. (2018). Infant statistical learning. *Annual Review of Psychology, 69*, 181−203.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*, 1926−1928.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language, 35*(4), 606−621.

Shatzman, K. B., & McQueen, J. M. (2006). Segment duration as a cue to word boundaries in spoken-word recognition. *Perception, & Psychophysics, 68*(1), 1−16.

Skoruppa, K., Nevins, A., Gillard, A., & Rosen, S. (2015). The role of vowel phonotactics in native speech segmentation. *Journal of Phonetics, 49*, 67−76.

Steinhauer, K., Alter, K., & Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience, 2*(2), 191−196.

Suomi, K., McQueen, J. M., & Cutler, A. (1997). Vowel harmony and speech segmentation in Finnish. *Journal of Memory and Language, 36*(3), 422−444.

Tremblay, A., & Spinelli, E. (2013). Segmenting liaison-initial words: The role of predictive dependencies. *Language and Cognitive Processes, 28*(8), 1093−1113.

White, L., Mattys, S. L., & Wiget, L. (2012). Segmentation cues in conversational speech: Robust semantics and fragile phonotactics. *Frontiers in Psychology, 3*, 375.

Woodrow, H. (1909). *A quantitative study of rhythm: The effect of variations in intensity, rate and duration.* New York: Science Press.

Word segmentation cues in the process of spoken language

YU Wenbo; LIANG Dandan

(*School of Chinese Language and Culture, Nanjing Normal University, Nanjing* 210097, *China*)

Abstract: Words are generally considered as the basic unit of language processing. Hence word segmentation is a vital step for language comprehension. In speech processing, cues for word segmentation may be phonological, grammatical or semantic. Phonological cues can be further classified as statistic, phonotactic and prosodic, while prosodic information involves stress, duration and pitch. Phonological cues are generally acquired at the initial stage of language learning, and they differ as the linguistic environment changes. Semantic and grammatical knowledge provide high-level cues which constrains word segmentation at later stage. It is suggested that future research focus on the trajectory of segmentation cues in a lifespan and the specificity of language in the process of word segmentation.

Key words: spoken language; word segmentation; phonology; semantics; grammar